Spatial analysis of road crashes involving vulnerable road users in support of road safety management strategies

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 45 (2020) 394–401

Transportation Research Procedia

www.elsevier.com/locate/procedia

AIIT 2nd International Congress on Transport Infrastructure and Systems in a changing world (TIS ROMA 2019), 23rd-24th September 2019, Rome, Italy

Spatial analysis of road crashes involving vulnerable road users in support of road safety management strategies

Marco Bassani[a]*, Leonardo Rossetti[a], Lorenzo Catani[a]

[a]*Politecnico di Torino, 24, corso Duca degli Abruzzi, Turin 10129, Italy*

Abstract

On urban road networks, approximately 2 out of 3 fatalities involve pedestrians, cyclists and motorcyclists, collectively referred to as "vulnerable road users" (VRU) due to their insufficient physical protection in the event of a collision. For a safer and more sustainable road transportation system, adequate protective countermeasures need to be introduced for this user category. However, related (and limited) resources restrict any safety improvements to certain high-risk sites with elevated rates of road traffic collisions. This study reports the results of a spatial distribution analysis of traffic collisions involving VRU in Turin over the period 2006-2016. The traffic road collisions database from the Italian National Institute of Statistics (ISTAT) was used for this purpose. Crash data were firstly geo-localized, and then analyzed using Geographic Information System technologies. A cluster analysis and a Kernel Density estimation were used to build spatial patterns of crashes involving VRU. Hazardous sites were identified on a metropolitan scale. Incorrect estimates of the actual collision frequency, which are typical of studies conducted over short periods, were avoided by considering only those sites where collision rates remained significantly high throughout the entire observation period (eleven years). The results show that clusters occur at intersections, many of which are located along corridors affected by heavy traffic flows and wide cross-sections. A further analysis was conducted to explain the role played by the geometric configuration (layout) of most hazardous sections and intersections in the level and severity of injuries and fatalities.

Keywords: road traffic collision; spatial analysis; vulnerable road users; Kernel density estimation; hazardous road locations; black hotspots.

* Corresponding author. Tel.: +39 011 564 5635.
E-mail address: marco.bassani@polito.it

10.1016/j.trpro.2020.03.031

1. Introduction

Road collisions have a social, personal and economic impact on society. They result in 1.24 million deaths each year, about 3,000 deaths per day (World Health Organization, 2015) and are the ninth leading cause of death worldwide (about 2.2% of global mortality), and the first for the 15-29 age group. Everyone circulating on the road is at risk, i.e. not only vehicle drivers and their passengers. Approximately 50% of deaths caused by road crashes involve vulnerable road users (i.e., pedestrians, cyclists and motorcyclists - VRU) due to their insufficient physical protection in the event of a collision with vehicles.

The basis for any decision on the most appropriate countermeasure lies in an analysis of past events in order to identify any critical points in the road network where a significant concentration of road crashes has occurred. These analyses are performed by using techniques that combine geolocation tools and spatial interaction models, now implemented in Geographic Information System (GIS) (Jayan and Ganeshkumar, 2010). A knowledge of the spatial distribution of crashes allows road safety specialists to investigate hazardous road locations (HRL) (Elvik, 2008; Mohaymany et al., 2013). Although methodologies for the identification of HRL are consolidated in literature, issues relating to availability of data, data quality, and data processing and treatment, arise frequently.

The paper addresses these issues and evaluates the effects of road network modification and implementation, the precision of recorded crash data, false positives and false negatives, and proposes new solutions. According to this methodology, HRL for VRU in the Turin (North-west of Italy) case study were identified for the 2006-2016 period.

2. Methodology

Collision events are records in a 2D space, which can be expressed in geographic terms (longitude, latitude), in cartographic (East, North) coordinates, or in the local plane (X, Y). GIS software facilitates the storage and processing of such georeferenced data (Aronoff, 1989). GIS represents and manages the large amount of information present in road collision databases, and provides tools for spatial data analysis, such as the identification of sites with a significant concentration of road collisions. These sites may be classified as HRL since the number of collisions considerably exceeds the average value observed in the surrounding area. The main goal of collision analysis is to determine whether the high rates recorded for certain sites are caused by specific conditions of the road environment or other factors.

The presence of clusters suggests a degree of spatial dependence between crashes; these concentrations can be caused by one or more defects of the infrastructure requiring a dedicated analysis. If positive, this spatial dependence, or spatial autocorrelation, indicates an aggregation of similar values of the investigated variable (Black, 1991). This means that two crashes occurring at or close to the same spot could have been caused by the same factor. From the identification of HRL and their defects, different countermeasures can then be designed and implemented to reduce crash frequency and/or severity.

According to Loo and Yao (2013), approaches to identify positive spatial autocorrelation may be divided into two major groups: (i) link-attribute approaches; and (ii) event-based approaches.

Link-attribute approaches provide for a segmentation of the road network into basic spatial units (BSU) and the counting of collision events within them; in this framework, crashes become attributes of these segments. However, these operations involve a significant computational effort, and entail different difficulties: (i) the correct choice of the BSU length, (ii) the impossibility of dividing different parts of a road network into segments of equal sized length, (iii) the difficulty with the interpretation of data from sections with different length, and (iv) the consequent loss of information.

Event-based approaches are divided into (i) distance-based methods, which define the presence of spatial aggregations of points, and in (ii) density-based methods, which instead identify such aggregations along the network. Among the distance-based spatial analysis methodologies, the Nearest Neighbor (NN) method (Clark and Evans, 1954) compares the characteristics of an observed set of distances between pairs of closest points ($\bar{d}_{obs}$):

$$\bar{d}_{obs} = \frac{1}{n}\sum_{i}^{n} d_{\min,i} \tag{1}$$

with distances that would be expected ($\bar{d}_{\exp}$) if points were randomly placed:

$$\bar{d}_{\exp} = \frac{1}{2}\left(\frac{n}{a}\right)^{-0.5} \qquad (2)$$

where $d_{min,i}$ is the i-th distance from the closest point, n is the number of observations, and a is the size of the study area. The NN index measures the similitude between observed and expected distances, by computing the difference (d), or the ratio (r) between $\bar{d}_{obs}$ and $\bar{d}_{\exp}$. Depending on the value of these two indexes, three different point patterns may be found. A clustered pattern ($d < 0$, $r < 1$) is frequently found in road collision data. If a collision pattern is more spread out, it exhibits a random spatial pattern ($d \rightarrow 0$, $r \rightarrow 1$). Although there may be some local clusters in this type of pattern, the overall structure is spread across the study area without any apparent pattern (this means that a road collision has an equal chance of occurring anywhere in the study area). The third type of pattern is uniform ($d > 0$, $r > 1$), and has rarely been found in road collision analyses.

The key tool for density based methods is the Kernel Density Estimation (KDE) which has been widely used in fields such as criminology (Anselin et al., 2000; Ummarino, 2013), epidemiology, biology, as well as in the study of fire trigger points (Amatulli et al., 2005), economic activities (Porta et al., 2009) and road safety (Shafabakhsh et al., 2017; Newaz et al., 2017). To avoid the incorrect identification of HRL, it is assumed that each single event affects the density in its spatial neighborhood, with a continuous, symmetrical, and decreasing probabilistic function through a factor of regression depending on the type of function itself and on the interpolation space around the same point (bandwidth). The density functions associated with each individual event are then cumulated to obtain the final density estimation (Levine, 2002). According to Fotheringham et al. (2000), the KDE can therefore be expressed for n points (events) according to:

$$\hat{f}(u,v) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} K\left(\frac{d_i}{h}\right) \qquad (3)$$

where $f(u, v)$ is the crash density estimate at the location (u, v), n is the number of observations, h is the bandwidth, K is the selected kernel function (e.g., normal, triangular, quadratic, quartic, etc.), and d_i is the distance between the location (u, v) and the location of the i-th observation. The two basic parameters are the K function and the bandwidth (h). The choice of the K function does not significantly affect the result (Loo and Anderson, 2015); in contrast, the value of the bandwidth significantly affects the output (see next section). Okabe et al. (2009), as well as Porta et al. (2009), suggest a range of values for h between 100 and 300 m with respect to urban areas, based on the average length of arcs in the road network. Thus, h strongly depends on the case study.

3. Crash database

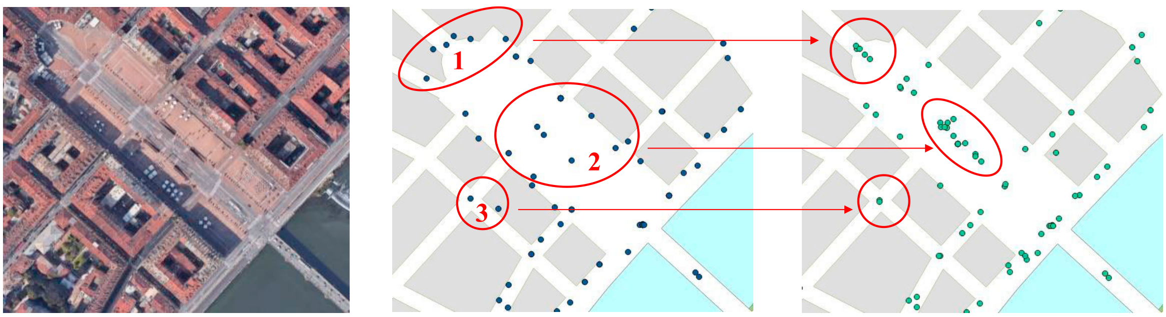
The starting source of data for the present study was the Regional collision database provided by ISTAT (National Institute of Statistics), related to the period 2006-2016; crash data for the Turin area was extracted from the database. Working from the definition of a road crash adopted by ISTAT (2015), the database lists all collisions involving at least one injury (Italy does not regard events resulting in property damage only as crashes). The data records were then divided into those involving “VRU” (i.e., where at least one VRU was involved) and vehicular crashes (called “noVRU”), consisting of 16,854 and 23,092 events respectively.

Since 2011, the ISTAT database has reported the geolocalization of crash events, while before 2011, the database provided the address and the closest house number for crashes which occurred along arcs, and the two street names for crashes which occurred at intersections. Discrepancies in data information in terms of the exact location of events are common in Italian Crash Databases (ISTAT, 2017), as well as in other countries (Mikulik and Hollo, 2007). In the case of data localized on the basis of street name and house number, the GPS Visualizer's (http://www.gpsvisualizer.com/) Address Locator tool was used to convert addresses into geographic coordinates. However, this tool failed in some cases, so a total of 1,632 inaccurate records out of 16,079 were identified and then manually corrected. It is worth highlighting that 775 VRU crash records in the official database reported incomplete geolocalization data, so they were excluded from further analyses and removed from the final database.

The effects of different localization methodologies are depicted in Fig. 1, where the same crash data in the period 2011-2016 are represented. Fig. 1B includes crash locations obtained from the closest street/square numbers information and then converted with the address locator tool, while Fig. 1C shows the same events localized using GPS position data included in the official record. Circles indicate three different data groups which have completely different spatial patterns. The reasons for a concentration on the west side (circle 1) and in the middle of the square (circle 2) is explained in Fig. 1A: collisions are concentrated along the most trafficked pedestrian crossing (circle 1) and along the pedestrian crossings close to the bus stops in the middle of Vittorio Veneto square (circle 2). Circle 3 shows the case of two events localized at the same point which are, however, wrongly located at two different points using the old geo-localization method. In the case of the largest square in the city, the effects of location inaccuracies are magnified, while smaller distances were observed along streets and ordinary intersections.

Road crashes are random events that fluctuate in time and space. Hence, the number of crashes at a specific location varies from month to month, and year to year. For the same period, the crash frequency varies at arcs and junctions. This variation over time for the same place reflects the "regression to the mean" phenomenon. In fact, it is highly probable that after a period with a high crash frequency recorded on a certain element (i.e., arc or intersection), a period with a lower frequency follows (AASHTO, 2010), and vice versa; i.e., data continuously regress towards the mean of the longer period. These fluctuations may determine false positive and false negative crash data aggregations. The false positive/negative issue can be curbed by aggregating data for periods of two or three consecutive years. The systematic presence of data aggregations for the same location reveals the presence of a true positive. Short periods lead to an increase in cases of false positives/negatives. However, long periods may include significant changes to the road network layout, which may have an effect on the crash distribution pattern. Therefore, any subdivision into periods must also take significant network transformations into account, in particular those which have the greatest impact on traffic operations and crash occurrence.

The six heat maps in Fig. 2 illustrate the case of crash clusters identified with the adoption of three different bandwidths equal to 50 (A, D), 100 (B, E) and 200 m (C, F) for the same case of Fig. 1. Fig. 2A, 2B and 2C show that an increase in bandwidth leads to the unification of distinct clusters, which renders the precise identification of the HRL problematic. In contrast, small values of *h* lead to the identification of numerous clusters concentrated in small areas. The choice of a large bandwidth value (200 m) would lead to the identification of a problem in a large area characterized by a single large cluster incorporating several road elements. Small values of *h* (50-100 m), instead, can serve to identify safety issues in specific HRL. In particular, Fig. 2A and Fig. 2B show two clusters, i.e., in the middle of the square (where there are bus stops) and in the area near the bridge (characterized by a high volume of pedestrian activity). With *h* equal to 200 m (Fig. 2C) these two aggregations become a single large cluster, thus obscuring the actual location of the criticalities. Similar operations may be carried out for the two-year period 2015-2016; with a value for *h* of 100 m, the aggregation in the center of the square, albeit less significant, is still present, while the cluster near the bridge is no longer there. This means that this point could be a false positive (in the case of an upward peak crash frequency in the period 2012-2014) or a false negative (in the event of a downward peak of the crash frequency in the period 2015-2016); in order to understand the nature of this site, it is necessary to extend the analysis over longer periods.



(A) (B) (C)

Fig. 1. (A) Aerial picture and (B, C) GIS crash localization around Vittorio Veneto square (Turin, Italy). Comparison between street/house number (B) and GPS (C) localization methods in the period 2011-2016.

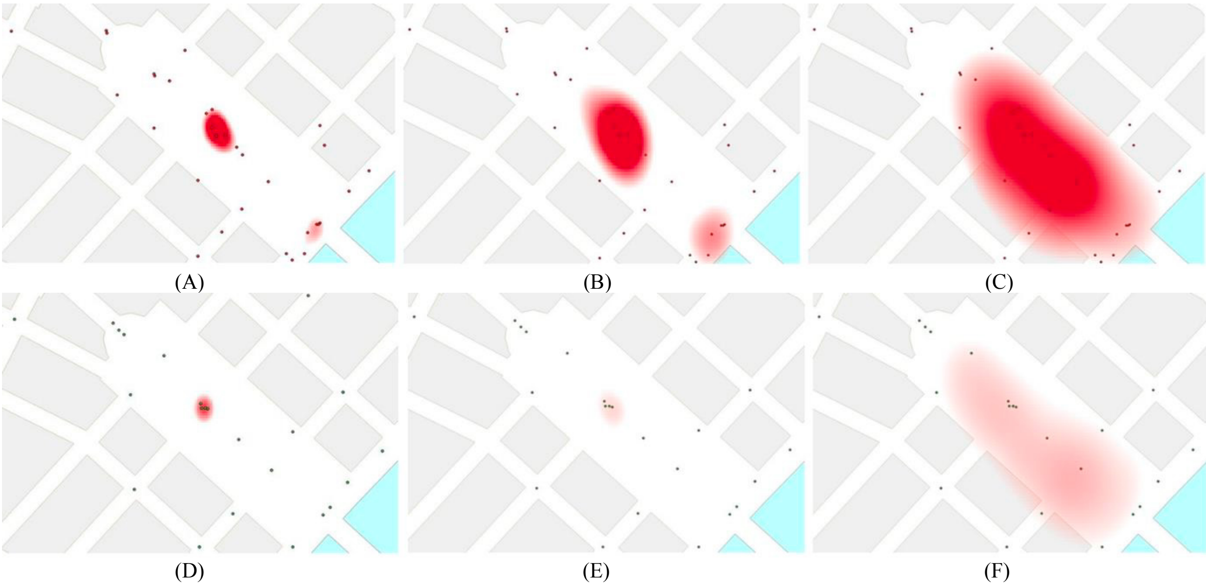


Fig. 2. Heat maps with h = 50 m (A, D), h = 100 m (B, E) and h = 200 m (C, F) in Vittorio Veneto square (Turin, Italy). Periods: 2012-2014 (A, B, C), and 2015-2016 (D, E, F). Red heat map indicates that the Kernel density is six standard deviations (SD) higher than the average in the study area.

4. Application of HRL methodology to the case study

Fig. 3A illustrates the distribution of 39,946 crashes which occurred in 2006-2016 in the city of Torino, differentiated by degree of severity (injury and fatalities). Although the database is affected by inhomogeneity issues regarding the crash data localization, which may lead to a wrong identification of crash aggregation, all eleven years were considered in the spatial distribution of road crashes. This decision was taken to maintain the informative content available (eleven-year period). Issues with cluster identification were partly addressed by the adoption of a bandwidth h of 100 m. Based on the most important road network modifications, the eleven-year period was divided into four periods of two years, and one of three years. Fig. 3 differentiates between crash events by severity (Fig. 3A), and type of road users involved in (Fig. 3B for VRU and Fig. 3C for "noVRU" crashes).

Fig. 3. (A) Fatal (379) and injury (39,567) crashes in Turin, Italy (total 39,946) in 2006-2016 (red points indicate fatalities, green points injuries); (B) VRU crashes (16,854) and (C) noVRU crashes (23,092) in the same period.

A preliminary evaluation of the points pattern, aimed at verifying the presence of a "clustered" structure was carried out. The NN indexes listed in Table 1 indicates that in the five study periods the ratio was constantly less than 1 in the range 0.387-0.437, i.e., the crash data assumed a clustered pattern. As a result, the spatial analysis will certainly lead to the identification of HRL. It was carried out assuming the Quartic Kernel function in eq. (3). The bandwidth (h) was determined by referring to the average arc length of the Torino road network (Table 2). It was estimated by analyzing the eight city districts separately. A value between half of the average arc length and the maximum arc length was assumed to avoid the incorporation within the same cluster of two different intersections. Excluding the old town (district 1), as well as districts 3 and 4, mean arc length values were found to be between 140 and 150 m. In relation to these results, a global scale bandwidth of 100 m, slightly above 50% of the average value characteristic of the arcs of peripheral areas, was adopted. This choice facilitated the differentiation between individual intersections with the exception of those in very close proximity to each other.

The analysis was carried out as per the QGIS software and related spatial analysis tools (https://www.qgis.org/). The output consisted of a continuous surface on Kernel Density from eq. (3), and heat maps were obtained by colouring the surface with individual colours depending on density intervals. Fig. 4 shows the example of heat maps for the 2015-2016 period for all recorded crashes (Fig. 4A), and those including at least one VRU (Fig. 4B). Fig. 4C includes the coloured scale with yellow, orange and red referring to the bands (2, 4 and 6 times the standard deviation – SD – above the mean – M – elevation of the density surface). For easier detection of most critical points on the network, the analysis only includes sites where the KDE function assumes density values higher than $M + 6 \cdot SD$. Data values for all five periods were considered.

Table 1. NN indexes for the study period 2006-2016 in Torino (Italy) according to eq. (1) and eq. (2).

Period (y-y)	2006-2007	2008-2009	2010-2011	2012-2014	2015-2016
NN (r) index	0.387	0.419	0.435	0.398	0.437

Table 2. Average length (in m) of Torino's arc network in the eight city districts.

District #	1	2	3	4	5	6	7	8
Average [m]	94.0	141.4	110.1	106.7	140.3	148.2	151.8	140.2

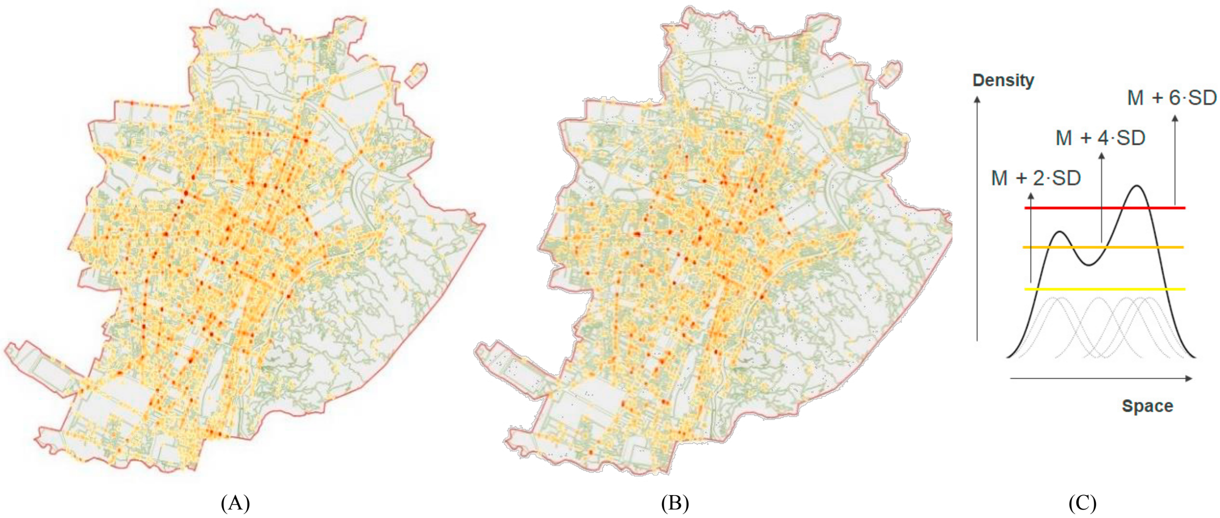


Fig. 4. Heat map of 6,176 crash events occurring in the 2015-2016 period (A), and heat map of 2,773 VRU crashes in the same period (B); (C) Color scale adopted for the two heat maps.

Fig. 5 shows the heat map resulting from the overlapping of density peaks (HRL for VRU with density larger than M + 6 · SD) that occurred in at least four of the five periods listed in Table 1. It shows how the criticalities are concentrated at the intersections along some of the main axes of the city, in particular: *corso Vittorio Emanuele II*, *corso Novara*, and the corridor constituted by *corso Lecce*, *corso Trapani* and *corso Siracusa* HRL are dispersed along the main urban trafficked routes. Intense traffic flows give rise to numerous conflicts between vehicles and VRU. In addition, wide carriageways with more than two lanes per direction expose pedestrians and cyclists to a higher rate of unsafe interactions with vehicles, as the former require more time to cross the intersection. Finally, along such corridors higher vehicular speeds are favored, and the higher speed differential between "noVRU" and VRU potentially leads to more severe collisions. Spatial analysis makes it possible to identify locations with a significant concentration of crashes; the actual causes for these concentrations need to be identified as per on-site monitoring and inspections.

5. Conclusions

To protect VRU it is essential to have safe infrastructure. VRU safety is paramount if we wish to promote non-vehicle mobility. As evidenced in this study, crashes involving at least one VRU account for about 40% of the total recorded, a significantly high share indeed. It is therefore necessary to carefully identify HRL in order to protect pedestrians, bikers and motorcyclists if we wish to promote this mobility component in urban areas.

The identification of HRL requires a careful evaluation of the study period to exclude cases of false positives and negatives, and to avoid the under- or overestimation of crash frequencies due to the "regression to the mean" phenomenon. The results show that critical VRU crashes occurred at the intersections of the main city avenues due to the higher number of conflicts with vehicular users. Specifically, the main hot spots are located along the first inner ring road of the city (*corso Siracusa*, *corso Trapani*, *corso Lecce, piazza Rivoli*), and along *corso Vittorio Emanuele II*. These sections are characterized by three main criticalities: (i) the high traffic flow along the main (central) carriageway, and along the lateral service carriageways; (ii) the speed difference between road users that traversed these sections; and (iii) the geometry of the cross section, which influences the level of exposure (time required to cross road) to danger of the VRU.

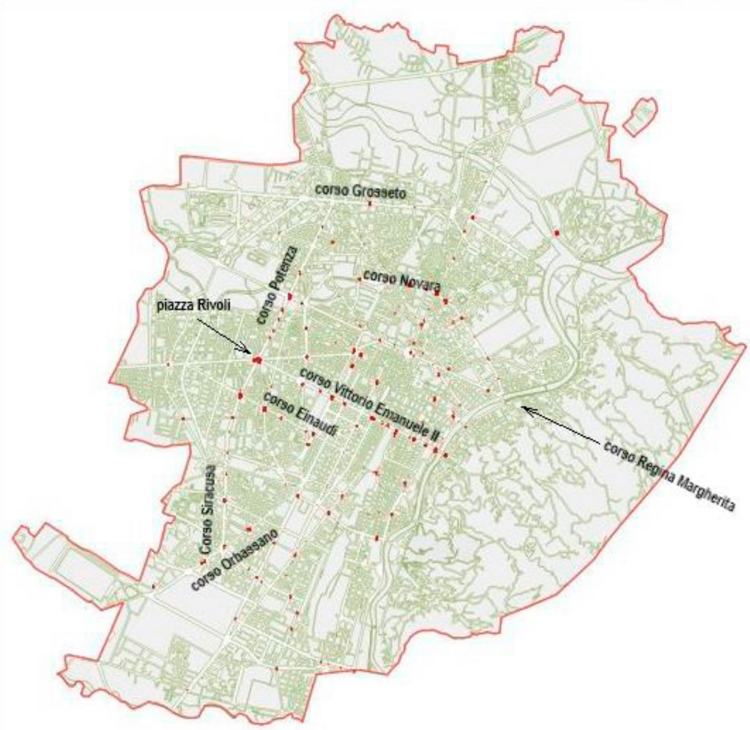


Fig. 5. HRL for vulnerable road users, in Turin (2006-2016).

The spatial analysis highlights the correlation between collisions and network structure. For the implementation of effective policies to contrast the rate of crashes, it is necessary to link these results to successive actions. Spatial analysis offers the opportunity to provide the analyst with an order of priorities. The local analysis of the HRL and the design of specific countermeasures can thus be addressed to increase the overall safety of VRU as well as other road user categories. It should be noted that the identification of HRL requires further investigation (site inspections) to capture any defects that will assist the analyst with the adoption of the most appropriate countermeasures, thus they need to be accurately identified.

References

American Association of State Highway and Transportation Officials (AASHTO), 2010. Highway Safety Manual. Washington, D.C.

Amatulli, G., Lovreglio, R., Leone, V., 2005. Geo-statistical Analysis for Pattern Evaluation of Fire Ignition Points in Rural Areas: The Adaptive Kernel Density Estimation Approach and its Implementation in the Territory of the Gargano National Park. Italian Journal of Forest and Mountain Environments 60.1, 85–97.

Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G., 2000. Spatial Analyses of Crime. Criminal justice 4.2, 213–262.

Aronoff, S., 1989. Geographic Information Systems: A Management Perspective. WDL publications.

Black, W. R., 1991. Highway Accidents: A Spatial and Temporal Analysis. Transportation Research Record 1318, 75–82.

Clark, P. J., Evans, F. C., 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35.4, 445–453.

Elvik, R., 2008. A Survey of Operational Definitions of Hazardous Road Locations in Some European Countries. Accident Analysis & Prevention 40.6, 1830–1835.

Fotheringham, A. S., Brunsdon, C., Charlton, M., 2000. Quantitative geography: perspectives on spatial data analysis. Sage.

Gutierrez-Osuna, R., 2004. Kernel density estimation. Available online: research.cs.tamu.edu/prism/lectures/pr/pr_l7.pdf.

ISTAT, 2015. Methodological notes for road accident analysis and report. Available online: https://www.istat.it/it/files//2016/11/nota-metodologica_incidenti-stradali_anno2015.pdf

ISTAT, 2017. Italian road accident database. Available online: https://www4.istat.it/it/archivio/incidenti+stradali/

Jayan, K. D., Ganeshkumar, B., 2010. Identification of Accident Hot Spots: A GIS Based Implementation for Kannur District, Kerala. International Journal of Geomatics and Geosciences 1.1, 51–59.

Levine N., 2002. CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0). Ned Levine and Associates Annandale, VA and The National Institute of Justice Washington, DC.

Loo, B. P., Yao, S., 2013. The Identification of Traffic Crash Hot Zones Under the Link-attribute and Event-based Approaches in a Network-constrained Environment. Computers, Environment and Urban Systems 41, 249–261.

Loo, B. P., Anderson, T. K., 2015. Spatial Analysis Methods of Road Traffic Collisions. CRC Press, Boca Raton, pp. 322.

Mikulik, J., Hollo, P., 2007. Road Accident Investigation Guidelines for Road Engineers. World Road Association PIRAC Technical Committee.

Mohaymany, A. S., Shahri, M., Mirbagheri, B., 2013. GIS-based Method for Detecting High-crash-risk Road Segments using Network Kernel Density Estimation. Geo-spatial Information Science 16.2, 113–119.

Newaz, K. M. S., Hasanat-E-Rabbi, S., Miaji, S., 2017. Spatio-temporal study of road traffic crash on a national highway of Bangladesh, 4th International Conference on Transportation Information and Safety (ICTIS). Alberta, Canada.

Okabe, A., Satoh, T., Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science 23.1, 7–32.

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., Wang, F., Scellato, S., 2009. Street Centrality and Densities of Retail and Services in Bologna, Italy. Environment and Planning B: Planning and design 36.3, 450–465.

Scheid, S., 2004. Introduction to kernel smoothing. Talk.

Shafabakhsh, G. A., Famili, A., Bahadori, M. S., 2017. GIS-based spatial analysis of urban traffic accidents: Case study in Mashhad, Iran. Journal of traffic and transportation engineering (English edition) 4.3, 290–299.

Toroyan, T., Iaych, K., Peden, M., 2015. Global Status Report on Road Safety 2015. Geneva: World Health Organization.

Ummarino, A. 2013. Una Introduzione ai Software per il Crime Mapping (in Italian). Rivista di Criminologia, Vittimologia e Sicurezza 7.1, 147–159.

World Health Organization, 2015. Global status report on road safety 2015.